

PROPERTY OF  
DETON UNIVERSITY LIBRARY

RECEIVED OCT 23 1939

**The  
PSYCHOLOGICAL  
RECORD . . .**

AUGUST, 1939  
Vol. III No. 14

**SOME FACTORS INFLUENCING THE DISTRIBUTION  
OF ASSOCIATED WORDS**

STUART W. COOK and B. F. SKINNER



**THE PRINCPIA PRESS, INC.**  
**BLOOMINGTON, INDIANA**

*Price of this number, 15 cents*

## SOME FACTORS INFLUENCING THE DISTRIBUTION OF ASSOCIATED WORDS<sup>1</sup>\*

By STUART W. COOK and B. F. SKINNER  
*University of Minnesota*

In the usual experiment on word association many different responses are elicited when a single stimulus word is presented to a group of subjects. The frequencies of the various words obtained from a large group are not unrelated. In the data collected by Kent and Rosanoff the most frequent response constituted on the average 25.8% of all the responses to the stimulus word, the second most frequent, 12.1%, and so on. When these per cent frequencies are plotted against rank on logarithmic paper, an approximately straight line is obtained (3). The present study is an analysis of similar data obtained by Schellenberg with a different list of stimulus words and a more homogeneous group of subjects (2). The relation between the frequencies and their ranks is examined, and in addition the effects of several factors which enter into the relation are estimated, including the frequency of the stimulus word in normal speech.

The 200 words in the Schellenberg list were drawn from various works on the subject of emotional complexes. In this respect they differed from the Kent-Rosanoff list, where a word was in general avoided if it was likely to call up a personal experience. The Schellenberg words were mimeographed in a booklet (described to his subjects as "A Speed Test—A Measure of Uncontrolled Association") in which space was provided for writing responses. The test was given to groups of entering students at the University of Minnesota in the fall of 1927. The 500 subjects who provided the part of the data used here were about equally divided between the sexes and had an average age of presumably about seventeen or eighteen years with a relatively small range.

The original tabulations obtained in this study have kindly been

<sup>1</sup> We are indebted to Dr. Evelyn Raskin for suggestions made during the planning of this study and to the N.Y.A. for clerical assistance in tabulating the data.

\* Recommended for publication by Dr. J. R. Kantor, June 30, 1939.

supplied by Professor Schellenberg. We have treated them in the following way. The frequencies of the words given in response to each stimulus word were arranged in order.<sup>1</sup> The mean frequency at each rank was then obtained and converted to a percentage by dividing by five hundred, the total number of responses obtained for each word. In Figure 1 these percentages are plotted against their rank on logarithmic paper. All points are plotted up to Rank 30; every fifth point is plotted thereafter up to Rank 100; and every tenth point thereafter. Similar percentages for the Kent-Rosanoff data, previously reported, are included for comparison.

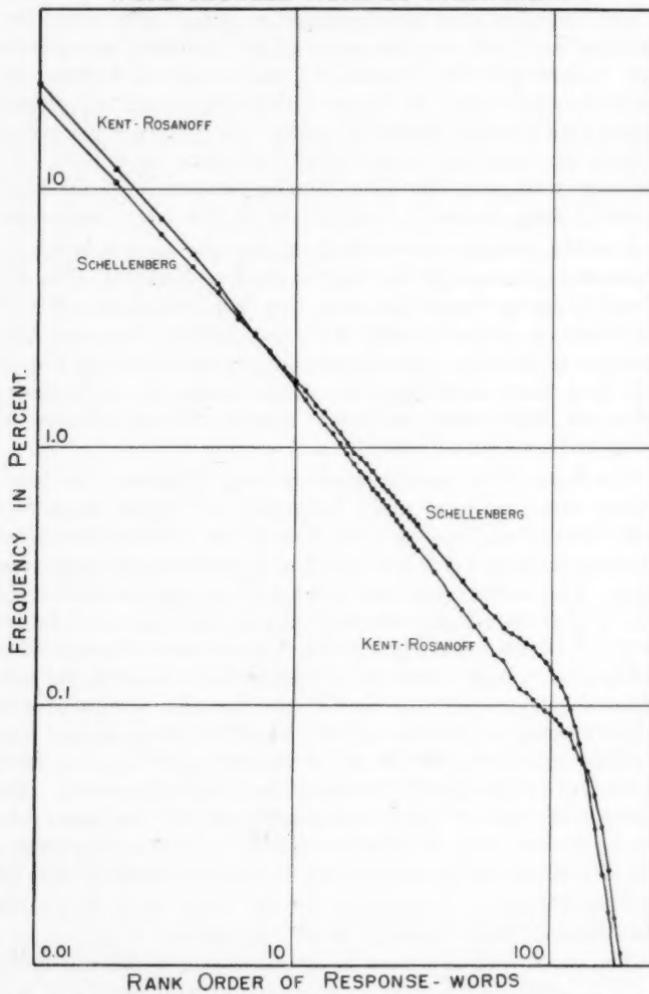
It will be seen that the Schellenberg data give an even better approximation to a straight line than do the Kent-Rosanoff. The difference in straightness is due to the first four frequencies, which in the latter case are low in terms of the main body of the curve. The first part of the curve depends upon the stimulus word and may be bent up or down according to the words chosen. As will be shown below, the Schellenberg words give a more balanced selection of strong and weak initial responses.

The slopes of the two curves are obviously different. The Schellenberg data yield lower initial frequencies and higher frequencies from Rank 10 on. In other words his subjects made less use of frequently associated words but agreed to a greater extent upon rarer words. This characteristic was found by Kent and Rosanoff to be true of their more highly educated subjects. In fact, the Schellenberg and Kent-Rosanoff groups stand in a relation to each other which almost exactly duplicates that between the educated and non-educated sub-groups in the Kent-Rosanoff study. Compared with subjects having a "common school" education, their subjects with a college education made less use of responses given by 1.5 or more per cent of the group and more use of less frequent responses. The curves in the present figure cross at approximately this point. Unless some other factor of selection was involved in their experiments, this difference may be assigned to a higher educational level in the Schellenberg group. Presumably the difference as it affects this experiment is one in the size of usable vocabulary.

The words used as stimuli in an experiment of this sort differ

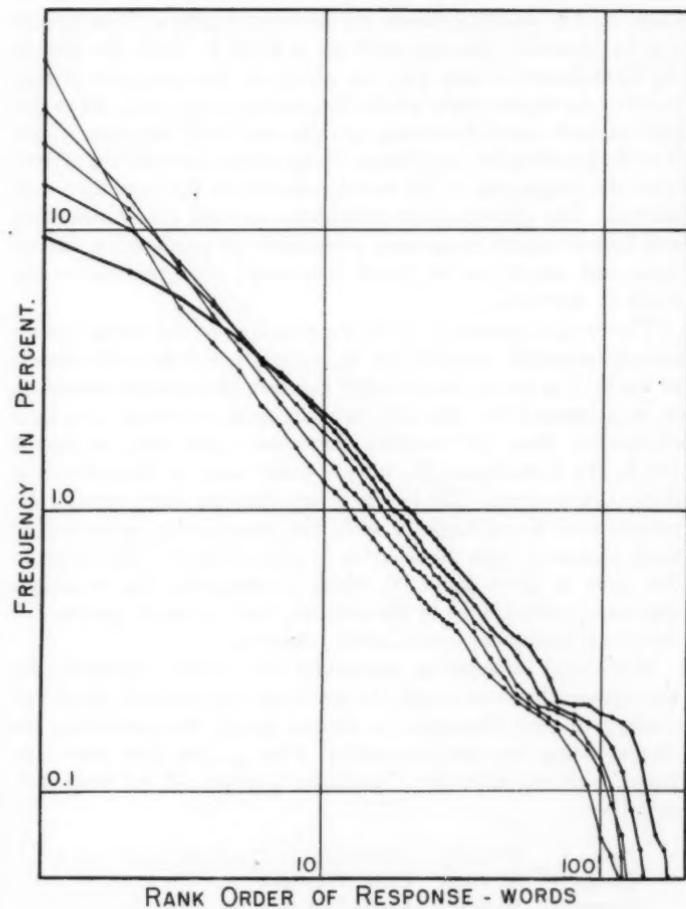
<sup>1</sup> A few entries of "no response" occasionally found in the Schellenberg tables were omitted.

FIGURE 1  
MEAN FREQUENCIES OF THE RESPONSES TO A STIMULUS-WORD PLOTTED AGAINST THEIR RANK



Data from experiments by Kent and Rosanoff and by Schellenberg are compared.

FIGURE 2  
FREQUENCIES OF RESPONSE-WORDS PLOTTED AGAINST  
RANK



The Schellenberg data have been classified according to the tendency of a stimulus-word to elicit one strong response.

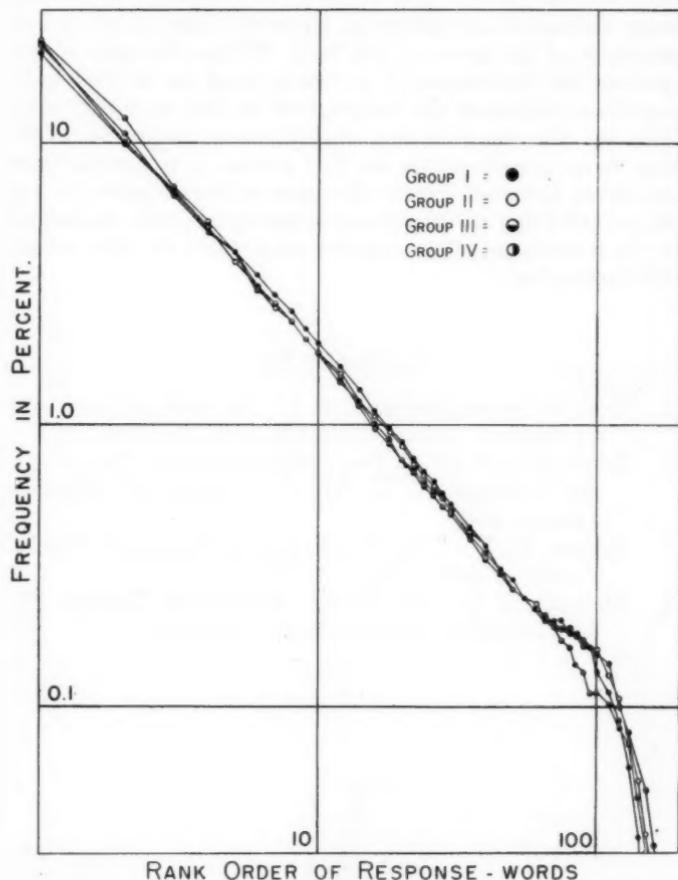
greatly in their tendency to elicit one very strongly associated response. In order to examine the effect of the strength of the most readily associated word upon the distribution as a whole the data have been divided into five groups according to frequency at Rank 1. The resulting curves are shown in Figure 2. The groups may be identified from the positions at Rank 1. As in the case of the Kent-Rosanoff data (3), the effect of the separation is confined to the higher ranks where the selection was made, the major part of each curve remaining straight and with the same slope. It is obvious that the straightness of any mean curve would depend upon the frequencies of the initial responses to the stimulus words selected. The difference in straightness between the Schellenberg and Kent-Rosanoff curves may presumably be attributed to differences with respect to the initial frequencies corresponding to the words in their lists.

The vertical separation of the main sections of the curves (which remain essentially parallel but in an inverted order with respect to Rank 1) is due to the different numbers of responses remaining to be accounted for after the most frequent responses have been eliminated. Since 250 stimulus words were used here, as against 100 in the Kent-Rosanoff study, a wider range of frequencies at Rank 1 is obtained. The lowest group (heaviest line) remains depressed until about Rank 10 while the lowest group in the earlier study recovered from the selection at about Rank 4. The hump in the curve at about Rank 70, which is apparently due to unique responses undetermined by the stimulus word, is again greatest for the group having the lowest initial frequency.

Additional information concerning the factors responsible for this distribution was sought by grouping the stimulus words according to their frequency in normal speech and combining the corresponding response frequencies. Four groups have been compared, with the aid of the Thorndike Teacher's Word Book (4), as follows:

Group I	70 words	Credit Numbers: 1a1, 1a2, 1a3, 1a4, 1a5.
Group II	63 words	Credit Numbers: 1b.
Group III	60 words	Credit Numbers: 2a, 2b.
Group IV	57 words	Credit Numbers: 3a, 3b, 4a, 4b, 5a, 5b, plus words not listed or given frequencies by Thorndike.

FIGURE 3  
FREQUENCIES OF RESPONSE WORDS PLOTTED AGAINST  
RANK



The Schellenberg data have been grouped according to the normal frequency of the stimulus-word as given in Thorndike's Teacher's Word Book. Group I contains the most common words; Group IV, the least common.

Group I contains the stimulus words found among the first hundred commonest words; Group II, words found between five hundred and a thousand; Group III, words between one and two thousand; and Group IV, words above two thousand. The resulting mean frequencies are plotted in Figure 3. There is no orderly separation of the curves on this basis. Within this range of frequencies, the commonness of a stimulus word has no statistically significant relation to the frequency of its first associated word. This was determined by comparing the mean frequencies of the most frequent responses for the four groups. It is apparent from the curves, moreover, that no subsequent difference arises. It may be concluded that the distribution of associated words is unrelated to the commonness of the stimulus word within the limits of the Schellenberg list.

#### REFERENCES

1. Kent, G. H., and Rosanoff, A. J. "A Study of Association in Insanity." *Amer. J. Insanity*, 1910, 67, 37-96.
2. Schellenberg, P. E. A Free Association Group Test for College Students. Ph. D. Thesis, University of Minnesota Library. 1929.
3. Skinner, B. F. "The Distribution of Associated Words." *Psychol. Record*, 1937, 1, 70-76.
4. Thorndike, E. L. *The Teacher's Word Book*. Teacher's College, Columbia University, 1921. 134 pages.

